Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at http://about.jstor.org/participate-jstor/individuals/early-journal-content.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

ON THE ACTUAL AND PROBABLE ERRORS OF INTERPOLATED VALUES DERIVED FROM NUMERICAL TABLES BY MEANS OF FIRST DIFFERENCES.

BY R. S. WOODWARD, C. E.

§ 1. All numerical calculations dependent on tabulated values of logarithms of numbers, sines, tangents etc., or on natural trigonometric functions are subject to certain errors arising from unavoidable inaccuracies in the tabular values themselves. The inaccuracies in the tabular values result from the neglect of figures beyond the tabular number of places, the tabular value being given always to the nearest unit in the last place. The actual error which may exist in the final result of any computation will depend on the form of the computation, or the manner in which the possible actual errors of the tabular values used are combined. The investigation of this class of errors becomes a matter of importance as well as interest whenever the highest degree of precision is sought in computations dependent on values derived from a given table. Thus, the practical question may arise, "What are the limits of accuracy attainable with a 5-place table of logarithms?" A complete discussion of the various problems that may arise in this connection would form an important branch of the theory of errors, and might well form the subject matter of a special treatise. In the present paper, however, it is proposed to consider only the most elementary of the problems, viz., that relative to the error which may exist in an interpolated value derived by means of first differences from consecutive tabular values. In other words, it is proposed to give some answer to the practical question, "What are the possible actual errors and what are the corresponding probable errors to which values interpolated from the ordinary 5-place or 7-place tables of logarithms are subject?"

Before taking up this problem, some remarks on the actual and probable errors of tabular values are required.

§ 2. The possible actual errors of tabular values are confined between the limits $+\frac{1}{2}$ and $-\frac{1}{2}$ of a unit in the last place of a tabular value, and all errors between these lemits are equally likely to occur. These facts may be taken as self-evident and as a necessary result of the method of constructing tables. The probable error, being defined as that error which is as likely to be exceeded as not, will be, in such a system of errors, one-half the maximum error, or $\pm\frac{1}{4}$ of a unit in the last place of a tabular value.

These statements may be readily verified in any special case. For example, if a large number of actual errors of tabular values of 5-place logari'ms be derived by means of a 7- or higher place table, it will be found that

1st, the numbers of positive and negative errors are sensibly equal;

2nd, the numbers of errors, without regard to sign, lying between equidistant limits are sensibly equal; and hence

3rd, one-half the errors will be greater and one-half less than $\frac{1}{4}$.

That the probable error in a system of errors of constant probability between given limits is one-half the max. or limiting error may also be proved in a more general manner, and some useful principles will result therefrom.

For this purpose, let ε be the general symbol for an error lying between the limits $\pm a$. The facility with which ε can occur may be denoted by $\varphi(\varepsilon)$ which must be in this case a constant, since by hypothesis all values of ε occur with the same facility. Hence we may write

$$\varphi(\varepsilon) = c, \text{ a constant.}$$

Supposing ε to vary continuously between the given limits we may make

$$\int_{-a}^{+a} \varphi(\varepsilon)d\varepsilon = c\int_{-a}^{+a} d\varepsilon = 1.$$

This is the analytical expression of the fact that the sum of all the probabilities of all possible errors within the given limits is unity. The value of the constant c results at once by integration. Thus

$$c\int_{-a}^{+a} d\varepsilon = 2ac = 1, \text{ whence}$$

$$c = \frac{1}{2a}. \tag{1}$$

Whatever form $\varphi(\varepsilon)$ may have, as long as $\varphi(+\varepsilon) = \varphi(-\varepsilon)$, or equal positive and negative errors are equally likely to occur, the probable error is the limit r in the definite integral

$$\int_{-a}^{-r} \varphi(\varepsilon)d\varepsilon = \int_{-r}^{0} \varphi(\varepsilon)d\varepsilon = \int_{0}^{+r} \varphi(\varepsilon)d\varepsilon = \int_{+r}^{+a} \varphi(\varepsilon)d\varepsilon = \tfrac{1}{4}\int_{-a}^{+a} \varphi(\varepsilon)d\varepsilon.$$

In the present case $\frac{1}{4}\int_{-a}^{+a} \varphi(\varepsilon)d\varepsilon = \frac{1}{4}$, and $\varphi(\varepsilon) = \frac{1}{2a}$. Hence

$$\int_{\pm r}^{0} \varphi(\varepsilon)d\varepsilon = \int_{\pm r}^{0} \frac{d\varepsilon}{2a} = \tfrac{1}{4}, \text{ whence } r = \frac{a}{2}.$$

§ 3. The errors to which interpolated values are subject may now be considered.

Let v_1 and v_2 be two consecutive tabular values and v' an interpolated value t tenths from v_1. Then the value of v' actually computed is

$$v' = v_1 + t(v_2 - v_1) = (1 - t)v_1 + tv_2. \tag{2}$$

Denoting the actual errors of v_1, v_2 and v' by ε_1, ε_2 and ε' respectively, we have

$$\varepsilon' = (1 - t)\varepsilon_1 + t\varepsilon_2. \tag{3}$$

The meaning of these equations should be clearly understood. Equation (2) means precisely what the symbols indicate, and does not imply any neglect of decimals beyond the tabular number. However, so far as errors here considered are concerned, it is practically sufficient to retain one figure in addition to the tabular number, retaining for example the 6th figure in interpolated values from a 5-place table and the 8th figure in interpolated values from a 7-place table. It will be noticed that $(v_2 - v_1)$ is the *tabular* difference. Equation (3) shows that the error ε' of v' cannot exceed $\pm\frac{1}{2}$, since ε_1 and ε_2 cannot surpass $\pm\frac{1}{2}$. An important property of ε_1 and ε_2 is their mutual independence; i. e., one of them cannot be expressed in terms of the other. They are also independent of t and continuous between the limits $\pm\frac{1}{2}$. Therefore ε' may have any value between the limits $+\frac{1}{2}$ and $-\frac{1}{2}$ for any value of t.

The species of interpolated value just defined is that which, as will presently appear, should be used always if the highest precision with a given table is desired. The species of interpolated value more frequently used, however, is that formed by abridging the product $t(v_2 - v_1)$ to the nearest unit in the last place of the tabular value. This neglect of decimals is a source of error additional to those arising from the errors of the tabular values, and may be denoted by ε_3. Calling the actual error of the interpolated value in this case ε'', the meaning of ε_1 and ε_2 being the same as defined above,

$$\varepsilon'' = (1 - t)\varepsilon_1 + t\varepsilon_2 + \varepsilon_3. \tag{4}$$

It will be observed that ε_3 in this equat'n differs from ε_1 and ε_2 in two important respects; viz., 1st, ε_3 is not independ't of t; 2nd, ε_3 is *discontinuous* for any value of t. Thus, suppose $t = \frac{1}{2}$; then the only possible values of ε_3 are 0, $+\frac{1}{2}$ and $-\frac{1}{2}$. Again, suppose $t = \frac{1}{3}$; then the only possible values of ε_3 are 0, $+\frac{1}{3}$ and $-\frac{1}{3}$. In a similar manner the possible values of ε_3 may be assigned for any value of t. The maximum possible value of ε'' varies, therefore, with t. A few cases may be written down.

For $t = \frac{1}{5}$ or $\frac{4}{5}$, the max. of $\varepsilon'' = \frac{4}{5}\times\frac{1}{2} + \frac{1}{5}\times\frac{1}{2} + \frac{2}{5} = \frac{9}{10}$.
" $t = \frac{1}{4}$ or $\frac{3}{4}$, " " " $\varepsilon'' = \frac{3}{4}\times\frac{1}{2} + \frac{1}{4}\times\frac{1}{2} + \frac{1}{2} = 1$.
" $t = \frac{1}{3}$ or $\frac{2}{3}$, " " " $\varepsilon'' = \frac{2}{3}\times\frac{1}{2} + \frac{1}{3}\times\frac{1}{2} + \frac{1}{3} = \frac{5}{6}$.
" $t = \frac{1}{2}$ " " " $\varepsilon'' = \frac{1}{2}\times\frac{1}{2} + \frac{1}{2}\times\frac{1}{2} + \frac{1}{2} = 1$.

It will also be noticed that for certain values of ε_3, ε'' will be necessarily positive or negative whatever values ε_1 and ε_2 may have. Thus for $t = \frac{1}{2}$, if the tabular difference $(v_2 - v_1)$ be odd, ε_3 must be either $+\frac{1}{2}$ or $-\frac{1}{2}$, and according as ε_3 is taken as $+\frac{1}{2}$ or $-\frac{1}{2}$ so will ε'' be either positive or negative, being confined between the limits 0 and $+1$ in the one case and 0

and — 1 in the other. On the other hand, the value of ε_3 may be such as to permit both positive and negative values of ε'' but with unequal limits for the two classes of errors. Thus for $t = \frac{1}{3}$, ε_3 must be either 0, $+\frac{1}{3}$ or $-\frac{1}{3}$. If $\varepsilon_3 = +\frac{1}{3}$, the possible negative values of ε'' must lie between 0 and $-\frac{1}{6}$, while the positive values of ε'' must lie between 0 and $+\frac{5}{6}$.

Since in a given numerical table the differences must be limited, the product $t(v_2-v_1)$ must vanish for $t=0$; and hence $\varepsilon_3 = 0$ for $t=0$.

Besides the two species of interpolated values already defined there may be two more, both of little practical importance, but of some theoretical interest from the fact that their definition gives some light on the real nature of the first two species. It is plain that instead of the tabular difference (v_2-v_1) there might be used the *true* differences, such as would be obtained practically by inserting in a given table differences derived from a higher place table and carried some decimals beyond the last place of values in the given table. Would there be any advantage or disadvantage in the use of such differences? There may be two cases. Let Δ be the true difference, v_1 the adjacent tabular value and t the interpolating factor. Then, supposing no figures omitted, the interpolated value becomes

$$v_1 + t\Delta. \tag{5}$$

The actual error of this value must be the same as the actual error of v_1, since there is no error in $t\Delta$. Denoting for the sake of distinctness the actual error of $v_1+t\Delta$ by ε''', we have

$$\varepsilon''' = \varepsilon_1.$$

If, however, the product $t\Delta$ be abridged to the nearest unit of the last tabular place, another source of error will be introduced. Calling the actual error from this source ε_4 and the actual error of the interpolated value ε^{iv}, there results

$$\varepsilon^{\text{iv}} = \varepsilon_1 + \varepsilon_4. \tag{6}$$

In this equation ε_4 may have any value between the limits $+\frac{1}{2}$ and $-\frac{1}{2}$ for any value of t greater than the reciprocal of twice the greatest difference in the table used.

For values of t greater than this limit ε^{iv} may have any value between the limits $+1$ and -1. For values of t less than the above limit, ε_4 must be less than $+\frac{1}{2}$, and hence ε^{iv} must be less than ± 1. When $t = 0$, $t\Delta$ vanishes and hence $\varepsilon_4 = 0$.

The characteristics of the four different species of interpolated values and the expressions for their actual errors may now be collected in tabular form for the purpose of ready comparison.

No. of species.	Obtained by means of	Expression for actual error.	Remarks.
1	Tabular diff.	$\varepsilon' = (1-t)\varepsilon_1 + t\varepsilon_2$	ε_1 and ε_2 are indep. of t and of each other & continu's between $\pm\frac{1}{2}$. ε_3 is dependent on t & discontinu's. ε_4 is con. bet. limits depen't on t.
2	" "	$\varepsilon'' = (1-t)\varepsilon_1 + t\varepsilon_2 + \varepsilon_3$	
3	True differ'e.	$\varepsilon''' = \varepsilon_1$	
4	" "	$\varepsilon^{\text{iv}} = \varepsilon_2 + \varepsilon_4$	

§ 4. Thus far the actual errors only of interpolated values have been considered. These errors may indeed be ascertained for any special case by a suitable computation, but such computation is generally inexpedient and of little interest. The probable error of each species of interpolated value is of interest and importance, however, since it affords a measure of the precision of such value. The probable error is, as already defined, that error which is as likely as not to be exceeded. It applies therefore to the aggregate of all possible errors in a system of errors, and may be taken as the representative error of the system. It applies in general only to continuous errors and is itself always a possible error of the system of errors.

This premised, let r be the general symbol for the probable error in a system of errors of which any one is ε, and let $\varphi(\varepsilon)$ express the law of facility of these errors. Then, supposing equal positive and negative errors of equal facility, or $\varphi(+\varepsilon) = \varphi(-\varepsilon)$, r is defined by the definite integral

$$\int_{-l}^{-r}\varphi(\varepsilon)d\varepsilon = \int_{-r}^{0}\varphi(\varepsilon)d\varepsilon = \int_{0}^{+r}\varphi(\varepsilon)d\varepsilon = \int_{+r}^{+l}\varphi(\varepsilon)d\varepsilon = \tfrac{1}{4}\int_{-l}^{+l}\varphi(\varepsilon)d\varepsilon,^{*} \quad (7)$$

in which $\pm l$ is the limiting value of ε. The determination of the probable error corresponding to ε', ε'', ε''' or ε^{iv}, requires therefore a knowledge of $\varphi(\varepsilon)$.

It will now be shown what forms $\varphi(\varepsilon)$ has for the system of errors represented by

$$\varepsilon' = (1-t)\varepsilon_1 + t\varepsilon_2.$$

Put $(1-t)\varepsilon_1 = x$ and $t\varepsilon_2 = y$. Then $\varepsilon' = x + y$, in which x may have any value between the limits $\pm(1-t)\frac{1}{2} = a$, say, and y may have any value between the limits $\pm t\frac{1}{2} = b$, say. All values of x and all values of y between these limits are equally likely to occur, which facts, as shown in § 2, are expressed analytically by

---

*This permits $\phi(\varepsilon)$ to be of such a form that $\phi(\varepsilon)d\varepsilon$ will represent a *relative* probability. If $\phi(\varepsilon)d\varepsilon$ represent an *absolute* probability then must

$$\int_{-l}^{+l}\phi(\varepsilon)d\varepsilon = 1.$$

$$\varphi(x) = \frac{1}{2a} \qquad \text{and} \qquad \varphi(y) = \frac{1}{2b}.$$

The probability of the occurrence of any particular x is

$$\varphi(x)dx = \frac{dx}{2a}.$$

Likewise the probability of the occurrence of any particular y is

$$\varphi(y)dy = \frac{dy}{2b}.$$

Since x and y are independent, the probability of their concurrence is

$$\varphi(x)\varphi(y)dx\,dy = \frac{dx\,dy}{4ab} = \frac{d\varepsilon\,dx}{4ab}, \text{ since } \varepsilon = x+y.$$

But since there must be in general an indefinite number of pairs of values of x and $y = \varepsilon - x$, which will produce ε, $\varphi(\varepsilon)d\varepsilon$ or the probability of ε must be equal to $dxd\varepsilon \div 4ab$ taken as many times as there are units in the range through which x may vary. In other words,

$$\varphi(\varepsilon)d\varepsilon = d\varepsilon \int \frac{dx}{4ab}.$$

In evaluating this integral, x must not surpass $\pm a$ and $\varepsilon - x = y$ must not surpass $\pm b$. For any value of ε lying between $-(a+b)$ and $-(a-b)$, assuming $a > b$, the limits of the integral are $-a$ and $(\varepsilon + b)$. This fact is rendered plain by a numerical example. Thus, suppose $a = \pm 5$ and $b = \pm 3$. Then $-(a+b) = -8$ and $-(a-b) = -2$. Suppose $\varepsilon = -6$ a number intermediate to -8 and -2. Then the following are the possible integer values of x and y which will produce $\varepsilon = -6$.

$$\begin{array}{rclcl}
\varepsilon' & & x \quad y & \qquad & \text{Limits of } x. \\
-6 & = & -5-1 & & -5 = -a \text{ and} \\
 & = & -4-2 & & -3 = \varepsilon'+b \\
 & = & -3-3 & &
\end{array}$$

Again, suppose ε' is -3, a number between the limits -8 and -2. Then

$$\begin{array}{rclcl}
\varepsilon' & & x \quad y & \qquad & \text{Limits of } x. \\
-3 & = & -5+2 & & -5 = -a. \\
 & = & -4+1 & & \\
 & = & -3+0 & & \\
 & = & -2-1 & & \\
 & = & -1-2 & & \\
 & = & 0-3 & & 0 = \varepsilon'+b.
\end{array}$$

Therefore, when ε lies between $-(a+b)$ and $-(a-b)$

$$\varphi(\varepsilon)d\varepsilon = d\varepsilon\int_{-a}^{\varepsilon+b}\frac{dx}{4ab} = \frac{\varepsilon+a+b}{4ab}d\varepsilon, \text{ or } \varphi(\varepsilon) = \frac{\varepsilon+a+b}{4ab}. \qquad (8)$$

Similarly, when ε lies between $-(a-b)$ and $+(a-b)$, the limiting values of x are $\varepsilon-b$ and $\varepsilon+b$. Therefore

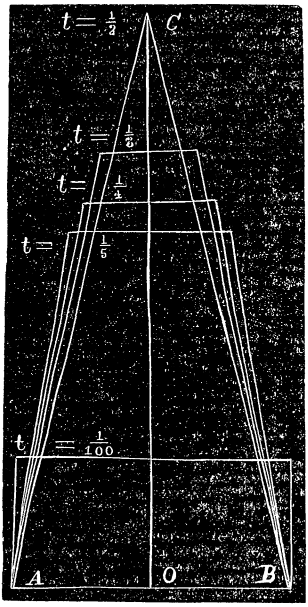
$$\varphi(\varepsilon)d\varepsilon = d\varepsilon\int_{\varepsilon-b}^{\varepsilon+b}\frac{dx}{4ab} = \frac{2b}{4ab}d\varepsilon, \quad \text{or } \varphi(\varepsilon) = \frac{2b}{4ab}. \qquad (9)$$

Finally, when ε lies between $(a-b)$ and $(a+b)$ the limits of x are $\varepsilon-b$ and $+a$, and hence

$$\varphi(\varepsilon) = \frac{-\varepsilon+a+b}{4ab}. \qquad (10)$$

The "probability curve" represented by equations (8), (9), (10), is made up of three straight lines. Equation (8) represents the line joining the p'ts whose coordinates are $-(a+b)$, 0 and $-(a-b)$, $(1\div 2a)$. Equation (9) represents the line joining the points whose coordinates are $-(a-b)$, $(1\div 2a)$ and $+(a-b)$, $(1\div 2a)$. Equation (10) represents the line joining the p'ts $+(a-b)$, $(1\div 2a)$ $+(a+b)$, 0. In the special case under consideration, $a = (1-t)\frac{1}{2}$ and $b = t.\frac{1}{2}$. Therefore

$$a + b = \tfrac{1}{2},$$
$$a - b = \tfrac{1}{2}(1-2t),$$
$$\frac{1}{2a} = \frac{1}{1-t}.$$



The figure shows the "probability curves" corresp'ing to several values of t. The origin is at O. AOB is the axis of ε and OC the axis of $\varphi(\varepsilon)$.

It may now be observed that eq's (8), (9), (10) give

$$\int_{-(a+b)}^{-(a-b)}\frac{\varepsilon+a+b}{4ab}d\varepsilon + \int_{-(a-b)}^{+(a-b)}\frac{2b}{4ab}d\varepsilon + \int_{+(a-b)}^{+(a+b)}\frac{-\varepsilon+a+b}{4ab}d\varepsilon = 1.$$

Therefore, by (7) the probable error in the system of errors defined by (8) - (10) is the limit r which will satisfy one of the two following conditions:

$$\int_{-(a+b)}^{-r}\frac{\varepsilon+a+b}{4ab}d\varepsilon = \int_{+r}^{+(a+b)}\frac{-\varepsilon+a+b}{4ab}d\varepsilon = \tfrac{1}{4}, \quad \ldots\ldots (\alpha)$$

$$\int_{-r}^{0}\frac{2b}{4ab}d\varepsilon = \int_{0}^{+r}\frac{2b}{4ab}d\varepsilon = \tfrac{1}{4}. \qquad (\beta)$$

[To be continued.]